International Journal of Advances in Intelligent Informatics
Vol. 2, No. 3, November 2016, pp. 149-156
ISSN: 2442-6571

Influence of overweight and obesity on the diabetes in the world on adult people using spatial regression

Tuti Purwaningsih [1,*], Baharudin Machmud [2]

Statistics Department, Universitas Islam Indonesia, Yogyakarta, Indonesia
*[1] tuti.purwaningsih@uii.ac.id *; [2] machmud8b@gmail.com*
** corresponding author*

ARTICLE INFO

Article history:
Received November 19, 2016
Revised November 25, 2016
Accepted November 25, 2016

Keywords:
Diabetes
Obesity
Overweight
Spatial model
SEM

ABSTRACT

This research discussed about the case of diabetes, overweight, and obesity which aimed to determine the factors that most affect the number of adult people with Diabetes from Obesity and Overweight in the world and looking for the best spatial model to make predictions in the next period. This research based on data WHO in 2015 from The 2016 Global Nutrition Report. At 5% level of significance for 2015, factor that influence diabetes is obesity and the most excellent spatial model used in the analysis is Spatial Error Model (SEM) that use Weight Level Order 1 and has R^2 value 81.82%.

I. Introduction

Diabetes, obesity, and overweight are now a staggering global burden and are approaching the scale of other forms of malnutrition. The prevalence of adult overweight, obesity, and diabetes is rising for every region and nearly every country. Based on data World Health Organization in the Global Nutrition Report 2015 [1] showed that an increasing number of people with Obesity, Overweight, and Diabetes in the world.

A regression modeling analysis to determine the factors that affect the rate of diabetes that influenced by the characteristic of region is important. In some cases, the response variable that observed has associated with observations in different areas, especially area that adjacent. The presence of spatial relationships in the response variable will cause estimation becomes inaccurate because the assumption of randomness an error is violated. To resolve these problems required a regression model that insert the spatial relationships between regions into the model. The presence of spatial relationships among area led to any spatial diversity into the model, so that the models were used is the regression model spatially.

Some methods have been developing are Spatial Autoregressive Model (SAR), Spatial Error Model (SEM), and Spatial Autoregressive Moving Average (SARMA). SAR, SEM, and SARMA are based on the effects of spatial lag and spatial error by using the approach area. The fundamental component from spatial model is spatial weighting matrix, this matrix reflects the relationship between a region with other regions [2]. In this research, the spatial weighting matrix that used is a spatial weighting Queen.

Based on the above, this research is solved using spatial regression with Spatial Autoregressive Model (SAR) and Spatial Error Model (SEM) to determine the factors that influence diabetes worldwide.

II. Literature Review

The regions are usually interrelated because of their proximity [3]. The similar performances in healthy field are usually attributed to their location. The spatial auto-correlation in the geographic approach is taken into account in the new healthy models used to forecast healthy issues. Therefore,

the spatial approach is not important only for analyzing the economic and social phenomena, but also for the policy decisions [4], such as looking for the factors that influencing the poverty status of districts in Java island for making government easier to prioritized the appropriate disctrict based on current condition [5].

According to the Indonesia's Endocrinologist Association (PERKENI) diagnostic criteria in 2015 [6], to suffer from diabetes if have fasting blood glucose levels > 126 mg/dL and at 2 hours after ate (postprandial) > 200 mg/dL. Blood glucose level varies for every individual every day where the number of blood glucose content will increase after the individual eat and will return to normal within 2 hours after eat. In normal condition, approximately 50% of glucose from food eaten will have perfect metabolism into carbon dioxide (CO2) and water, 10% into glycogen and 20% to 40% is converted into fat. In people with diabetes mellitus, all metabolic process disrupted by deficiency of insulin. The absorption of glucose into the cells decreased and the metabolism disturbed. This condition causes most of glucose remains in the blood circulation, causing hyperglycemia.

Obesity can be defined as excess body fat. The determinant that used is Body Mass Index (BMI). Whereas Overweight is stage before clinically obese. Based on WHO criteria [7], underweight was identified as BMI < 18.5 kg/m2, overweight as BMI 25.0-29.9 kg/m2, and obese as BMI ≥ 30.0 kg/m2.

According to WHO [7], body mass index (BMI) is the calculation of number from a person's weight and height. BMI obtained from the weight in kilograms divided by the square from height in meters (kg/m2). The value of the BMI in adults does not depend on age or sex. However, BMI may not correspond to the degree of obesity in different populations, due to differences in their body proportions.

III. Material and Methodology

The data that used in this research is secondary data that obtained from The 2016 Global Nutrition Report from WHO [8]. The data that obtained is data distribution of adult patients with Diabetes, Overweight, and obesity in 190 countries around the world in 2015. The data were collected through survey taken from each country.

In this research there are one dependent variable and two independent variables Diabetes (y), Overweight ($x_{overweight}$), and Obesity ($x_{obesity}$).

We started the research by looking for the classic regression model, the general model of spatial regression equation [9], [10] stated on (1) and (2).

$$Y = \rho WY + X\beta + \mu \quad (1)$$

$$\mu = \lambda W\mu + \varepsilon, \varepsilon \sim N(0, \sigma^2) \quad (2)$$

where Y is response variable matrix (n x 1), *X* for independen variable matrix (n x (p+1)), β for coefficients vector parameter regression (p+1)x1, spatial lag coefficient autoregression stated with ρ, λ for autoregression lag coefficient in error valued | λ | < 1, μ for vector error that assumed contain autocorrelation sized n x 1, ε for vector error sized nx1, normal distribution with mean zero and variance σ2I, W is spatial weighting matrix sized n x n, and number of observations / location stated with n.

There are four models that can be formed from the general model of spatial regression as follows: 1) If $\rho = 0, \lambda = 0$ then the equation number (1) is: $Y = X\beta + \varepsilon$ **(2)**, this equation is called spatial model Ordinary Least Square (OLS). If $\rho \neq 0, \lambda = 0$ then the equation number (1) is: $Y = \rho W\lambda + X\beta + \varepsilon$ **(3)**, this equation is called as regresion Spatial Lag Model (SLM) or Spatial Autoregressive Models (SAR). If $\rho = 0, \lambda \neq 0$ then the equation number (1) is: $Y = X\beta + \lambda W\mu + \varepsilon$ **(4)**, this equation is called as regresion Spatial Error Model (SEM). If $\rho \neq 0, \lambda \neq 0$ then the equation number (1) is: $Y = \rho W\lambda + X\beta + \mu, \mu = \lambda W\mu + \varepsilon$ **(5)**, this equation is called General Spatial Model or Spatial Autoregressive Moving Average (SARMA).

To determine the analysis, spatial dependencies test is required. Spatial dependencies tested with Lagrange Multiplier test [9]. The hypothesis Lagrange Multiplier test are:

(i) $H_0 : \rho = 0$ (no lag spatial dependencies)

$H_1 : \rho \neq 0$ (there is lag spatial dependencies)

(ii) $H_0 : \lambda = 0$ (no error spatial dependencies)
$H_1 : \lambda \neq 0$ (there is error spatial dependencies)

(iii) $H_0 : \rho, \lambda = 0$ (no lag and error spatial dependencies)
$H_1 : \rho, \lambda \neq 0$ (there are lag and error spatial dependencies)

Furthermore needed homogeneity test, spatial diversity using Pagan Breusch test [9], [11]. The hypothesis is:

$H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_n^2 = \sigma^2$ (among areas diversity / equal variances)

H_1 : minimal ada satu $\sigma_i^2 = \sigma^2$ (there are variations among areas / heteroskedasticity)

Perform spatial analysis needed spatial weighting matrix. Spatial weighting matrix basically is matrix that describes the relationship among area. In this research the spatial weighting matrix is Queen weighting matrix. Queen weighting matrix define $W_{ij} = 1$ for the area side by side or vertex met with area that be a concern, whereas $W_{ij} = 0$ to another area [10]. Spatial weighting matrix is the symmetric matrix and main diagonal are always zero.

There are several types of Spatial Weight (W): binary W, uniform W, invers distance W (non uniform weight) and and some W from real case of economics condition or transportation condition from the area. Binary weight matrix has values 0 and 1 in off-diagonal entries; uniform weight is determined by the number of sites surrounding a certain site in ℓ-th spatial order; and non-uniform weight gives unequal weight for different sites. The element of the uniform weight matrix is formulated as,

$$\text{Wij} = \begin{cases} \frac{1}{n_i^{(l)}} , j \; is \; neighbor \; of \; i \; in \; l-th \; order \\ 0 \qquad\qquad\qquad\qquad , others \end{cases} \tag{3}$$

n_i^((l)) is the number of neighbor locations with site-i in ℓ-th order. The non-uniform weight may become uniform weight when some conditions are met. One method in building non-uniform weight is based on inverse distance. The weight matrix of spatial lag k is based on the inverse weights 1/(1+d_ij) for sites i and j whose Euclidean distance dij lies within a fixed distance range, and otherwise is weight zero. Kernel Gaussian Weight follow this formulla :

$$w_j(i) = exp\left[-\frac{1}{2}\left({}^{d_{ij}}/_{b}\right)^2\right] \tag{4}$$

with d is distance between location i and j, then b is bandwith which is a parameter for smoothing function [5].

The models that have been obtained is then performed selecting best model. Selection of best model is done to get the most supportive factor of research. The criteria for selection of best model used is *Akaike's Information Criteria corrected (AICc)*.

$$AICc = AIC + (\frac{2p(p+1)}{n-p-1}) \tag{5}$$

IV. Results and Discussion

Fig. 1 shows the mapping that rates of diabetes most occurred in middle eastern areas. Based Conditional Map of Diabetes, Overweight, and Obesity (Fig. 2) in 190 Countries rates of Diabetes, Overweight, and Obesity many occur in the Northern Hemisphere and Australia. While at least in the region of Southeast Asia and each country in Africa.

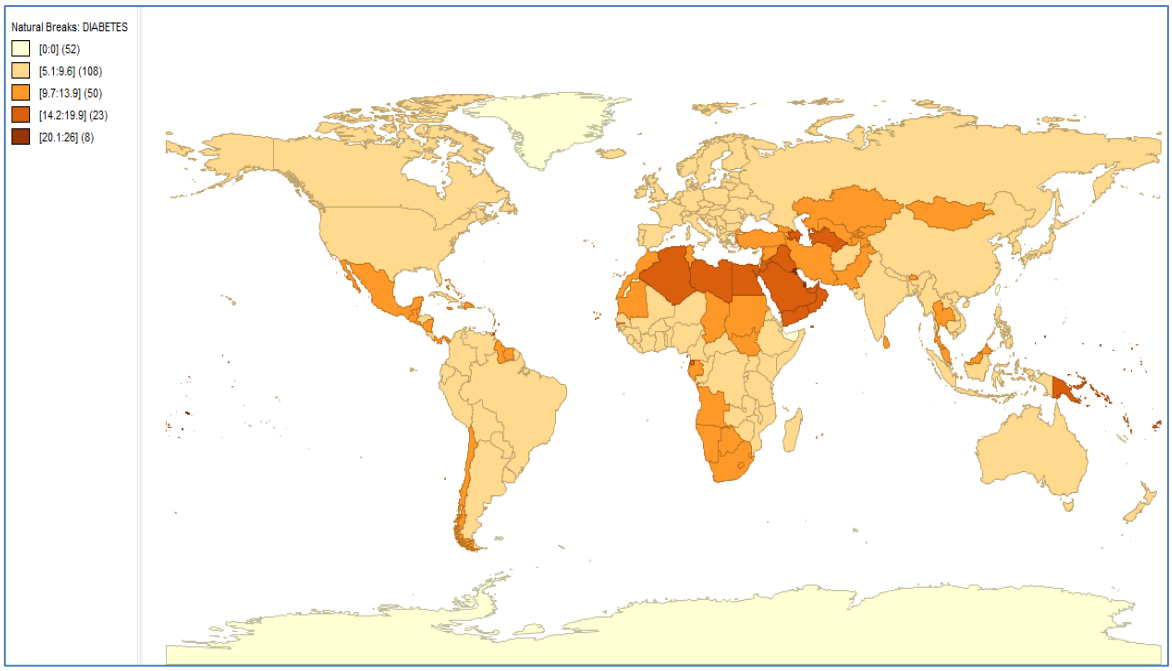

Fig. 1. Mapping Diabetes rate in 190 Countries

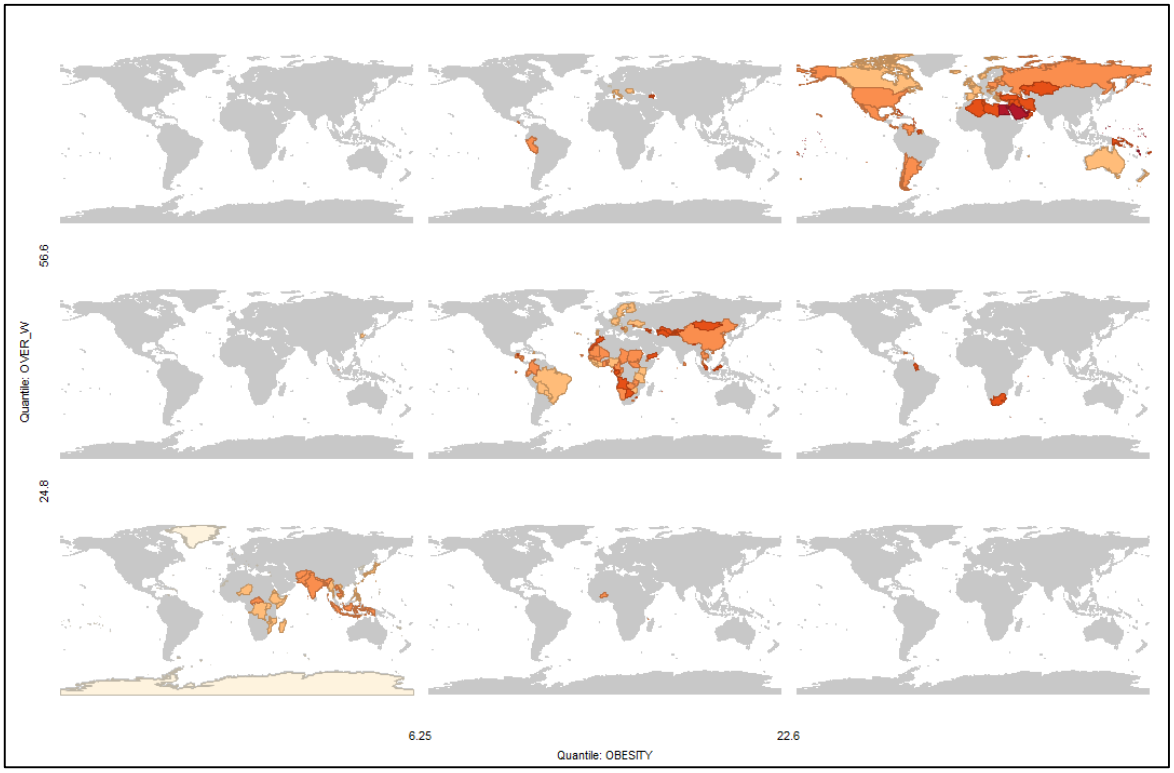

Fig. 2. Conditional Map of Diabetes, Overweight, and Obesity in 190 Countries

Moran's I index for 2013 is positive (0.156) and close to zero, indicating a positive spatial correlation (Fig. 3).

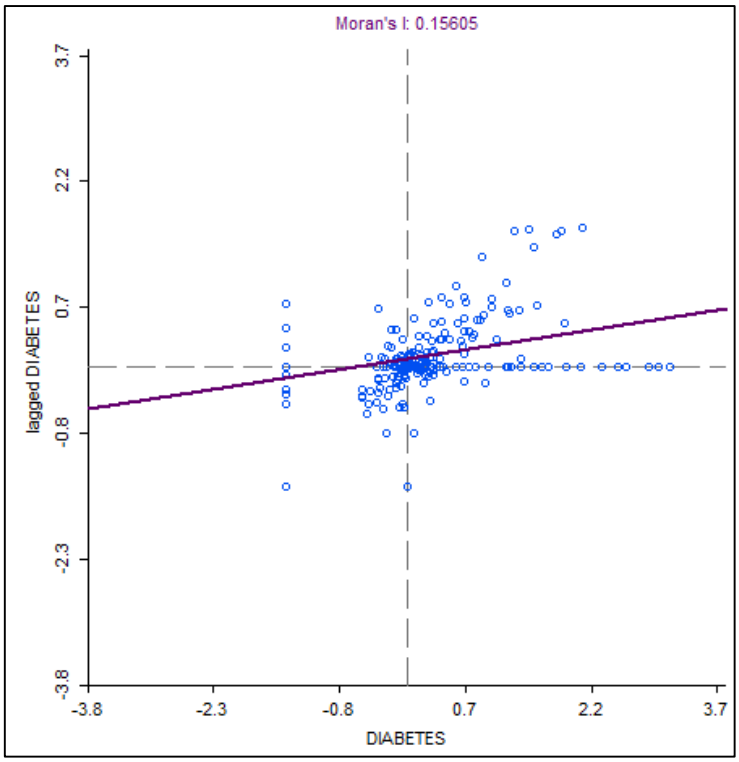

Fig. 3. Moran's I index of diabetes rate in 190 counties

However, a strong spatial auto-correlation can be observed between the diabetes rate in 2015 in 190 counties compared to the values for obesity and overweight. A bivariate Moran's I index of 0.514 for obesity (Fig. 4(a)) and 0.585 for overweight confirms a strong positive spatial autocorrelation (Fig. 4(b)).

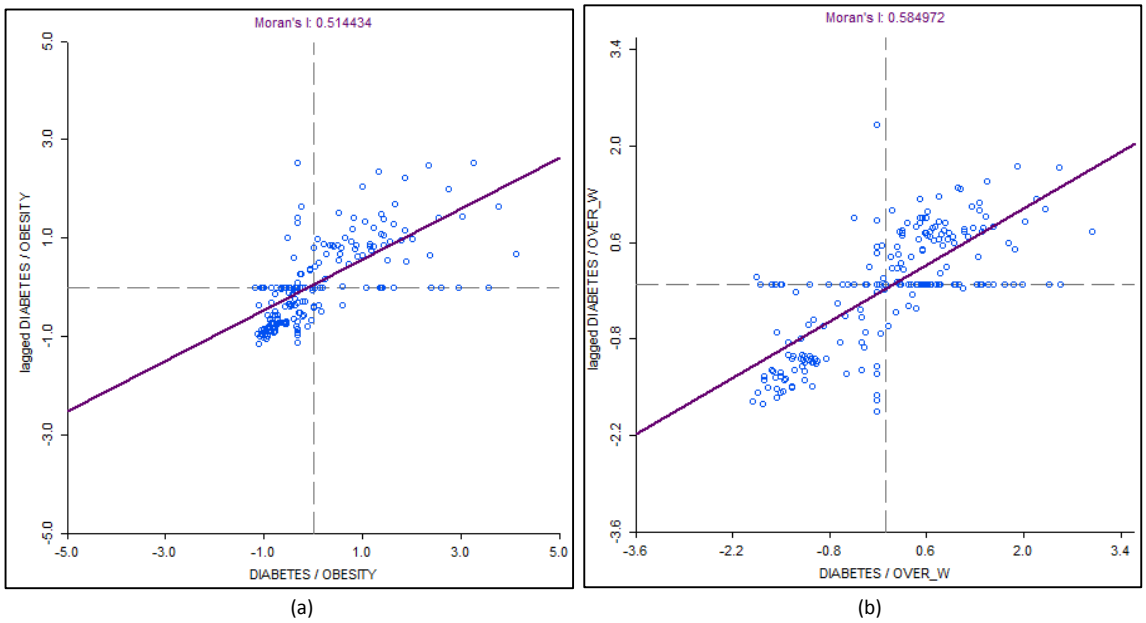

Fig. 4. Bivariate Moran's I index of diabetes, overweight, and obesity rate in 190 counties

Based on Table 1, there is one variable that have significant influence to the response variable (Diabetes) that is obesity. So variable that affects the number of adult people with diabetes from overweight and obesity is obesity.

Table 1. Identification Patterns of Relationships between Independent and Response Variables.

Variable	Coefficient	Probability
Constant	2.072	0.000
Over Weight	0.05768	0.062
Obesity	0.26685	0.000

Based on the classic regression analysis with order 1 obtained coefficient of determination (R^2) is 0,667 that means the regression model can explain 66,7% of the diversity total while the remaining 33,3% explained by other variables outside the model. Classical regression model that formed is

$$\bar{y} = 0.05768x_{\text{overweight}} + 0{,}26585x_{\text{obesity}} \quad (6)$$

Interpretation of the regression model is every increase of 1 unit on Obesity variable, it can increase the percentage of Diabetes (y) as much 0,265%. The first step to develop spatial model is initial identification model using LM (Lagrange Multiplier) (Tabel 2).

Table 2. Result of initial identification spatial dependency

Spatial dependency test	Value	P-*Value*
Lagrange Multiplier (lag)	10.783	0.001
Lagrange Multiplier (*error*)	1.#INF	0.000

A. Identification of dependency lag

Based on the information in Table 2 can be seen that P-value LM lag as much 0,001 by using α = 5%. This means that there is lag spatial dependencies that need to be continued to make *Spatial Autoregressive Model* (SAR). Table 3 shows the results of SAR parameter estimation.

Table 3. SAR parameter estimation

Variable	Coeff	Z	P-Value
W_Diabetes	0.140	3.205	0.001
Constant	1.818	4.451	0.000
Over_W	0.011	0.034	0.742
Obesity	0.344	0.066	0.000

SAR model that obtained as follows

$$\bar{y}_i = 0.05768x_{\text{overweight}} + 0{,}26585x_{\text{obesity}} + \varepsilon_i \quad (7)$$

$$\varepsilon_i = 0.140 \sum_{i=1, i \neq j}^{n} W_{ij} Y_j \quad (8)$$

Based on calculations using GeoDa obtained $R^2 = 0.681$ it means that model can explain variation from Diabetes as much 68,10% and the remaining 31,90% explained by other variables outside the model.

B. Identification of dependency error

Based on the information in Table 2 can be seen that P-value LM error as much 0.000 by using α = 5%. This means that there is error spatial dependencies that need to be continued to make *Spatial Error Model* (SEM). The results of SEM parameter estimation shown in Table 4.

Table 4. SEM parameter estimation

Variable	Coeff	Z	P-Value
Lambda	0.693	13.971	0.000
Constant	1.177	3.599	0.000
Over_W	0.098	3.585	0.000
Obesity	0.239	4.330	0.000

SAR model that obtained as follows

$$\bar{y}_i = 0.098x_{\text{overweight}} + 0{,}239x_{\text{obesity}} + \mu_i \quad (9)$$

$$\mu_i = 0.693 \sum_{i=1, i \neq j}^{n} W_{ij} Y_j + \varepsilon_i \quad (10)$$

Based on calculations using GeoDa obtained $R^2 = 0.8182$ it means that model can explain variation from Diabetes as much 88,82% and the remaining 11,18% explained by other variables outside the model.

C. Comparison of AICc and R^2 value

After obtained some models, the best model selection needs to be done. Selection of the best model using AICc and R^2 value criteria. A model can be concluded that the model is good model when the AICc value is small and R^2 value is big. Based on Table 5 obtained information that the model *Spatial Error Model* (SEM) Order 1 is the best regression model.

Table 5. Comparison of AICc and R^2 value from Model

Order 1			**Order 2**		
Model	***AICc***	***R^2***	***Model***	***AICc***	***R^2***
Classic Regressio	1260.75	0.6676	Classic Regressio	1260.75	0.6676
SAR	1253.96	0.6809	SAR	1254.40	0.6796
SEM	1146.64	0.8182	SEM	1177.62	0.7795

V. Conclusion

Based on the results and discussion, can be concluded following matters:

1. The factor that most influence the number of adult people with Diabetes from Overweight and Obesity is Obesity.
2. The best spatial regression model that predicted Diabetes in the world is *Spatial Error Model* (SEM) that using *Weight Level* order 1 which has R^2 81,82% with model as in (9) and (10).

References

[1] L. J. Haddad et al., Global Nutrition Report 2015: Actions and accountability to advance nutrition and sustainable development. Intl Food Policy Res Inst, 2015.

[2] R. Dubin, "Spatial weights," Sage Handb. Spat. Anal., pp. 125–158, 2009.

[3] M. D. Ward and K. S. Gleditsch, Spatial regression models, vol. 155. Sage, 2008.

[4] D. J. Lacombe, "Does econometric methodology matter? An analysis of public policy using spatial econometric techniques," Geogr. Anal., vol. 36, no. 2, pp. 105–118, 2004.

[5] A. S. Fotheringham, C. Brunsdon, and M. Charlton, Geographically weighted regression: the analysis of spatially varying relationships. John Wiley & Sons, 2003.

[6] S. A. Soelistijo et al., Konsensus pengelolaan dan pencegahan diabetes melitus tipe 2 di Indonesia 2015. PB. PERKENI, 2015.

[7] W. H. Organization, Obesity: preventing and managing the global epidemic, no. 894. World Health Organization, 2000.

[8] W. H. Organization and others, Global report on diabetes. World Health Organization, 2016.

[9] L. Anselin, "Spatial Econometrics: Methods and Models. Dordrecht: Kluwer Academic Publishers.," 1988.

[10] J. P. LeSage, "The theory and practice of spatial econometrics," Univ. Toledo. Toledo, Ohio, vol. 28, p. 33, 1999.

[11] R. P. Haining, Spatial data analysis: theory and practice. Cambridge University Press, 2003.